RP NGOMA COLLEGE

DATA MING AND WARE HOUSE

ETL WORK

Group Members:

|  |  |  |
| --- | --- | --- |
| **REG NUMBER** | **FIRST NAME** | **LAST NAME** |
| 23RP00470 | Jean Baptiste | NDUWAYEZU |
| 24RP08869 | Claudine | KWIZERA |

Store Books ETL and Data Warehouse

Project Report

Project Overview:

This project aims to transform the existing Store Books Sales database into a robust data warehouse, enabling advanced analytics and reporting capabilities. The project involves the use of Extract, Transform, Load (ETL) processes to migrate and transform data, and the implementation of SQL Server Analysis Services (SSAS) to create a cube for in-depth data analysis.

**Tools and Technologies Used**

* **Database Management Systems (DBMS)**
* **MySQL database:** for designing and managing the data warehouse.
* **SQL Server Management Studio (SSMS)**: for database management and query execution.
* **SQL Queries**: for retrieving data from the **source database** and load them into Datawarehouse.
* **Data Modeling & Schema Design**
* **Star Schema Design**: Used as data warehouse structure.
* **Python**: Data processing and transformation.

**How the Store Books Data Warehouse Project implemented**

Implement of **Store Books Data Warehouse Project** involves, **ETL (Extract, Transform, Load)** process which is a critical part of creating a data warehouse.

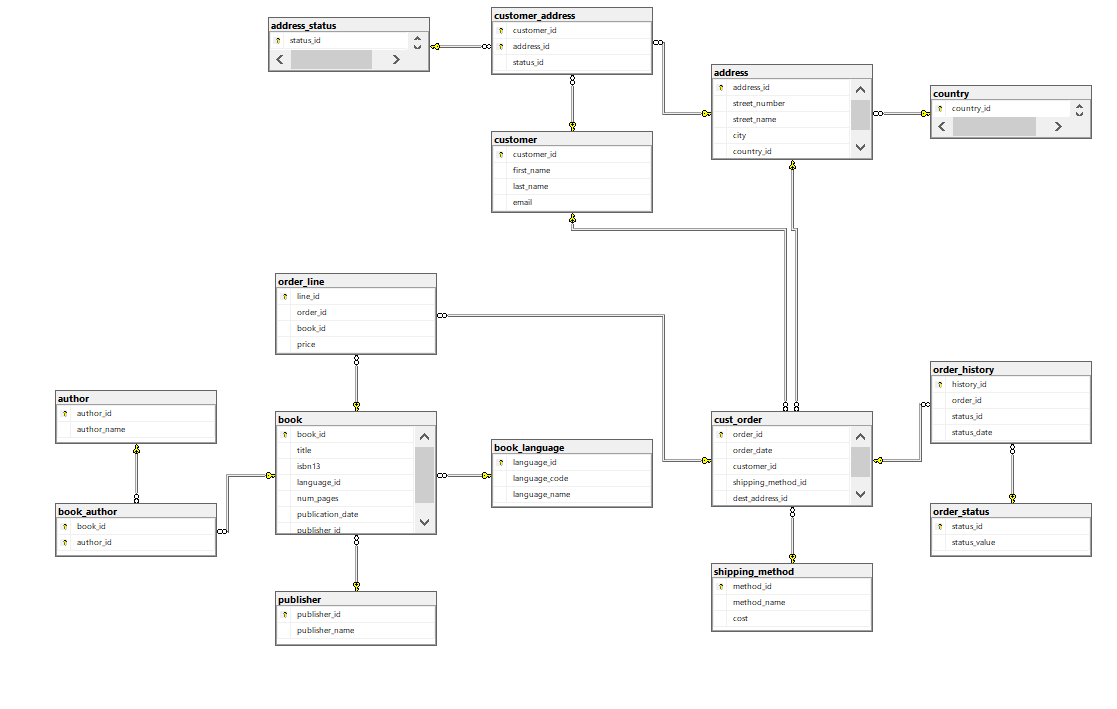
**ETL** involves:

* Extracting data from the source system,
* Transforming data to fit the warehouse schema,
* Loading data into the data warehouse.

1. **Extracting data from the source system**

The Extract phase focuses on retrieving data from the **gravity\_books** database. This step ensures that the necessary raw data is gathered for further processing and transformation.

Figure 1: Data source schema



Data from Source Tables Database (gravity\_books)

**Import necessary libraries and install some missed**

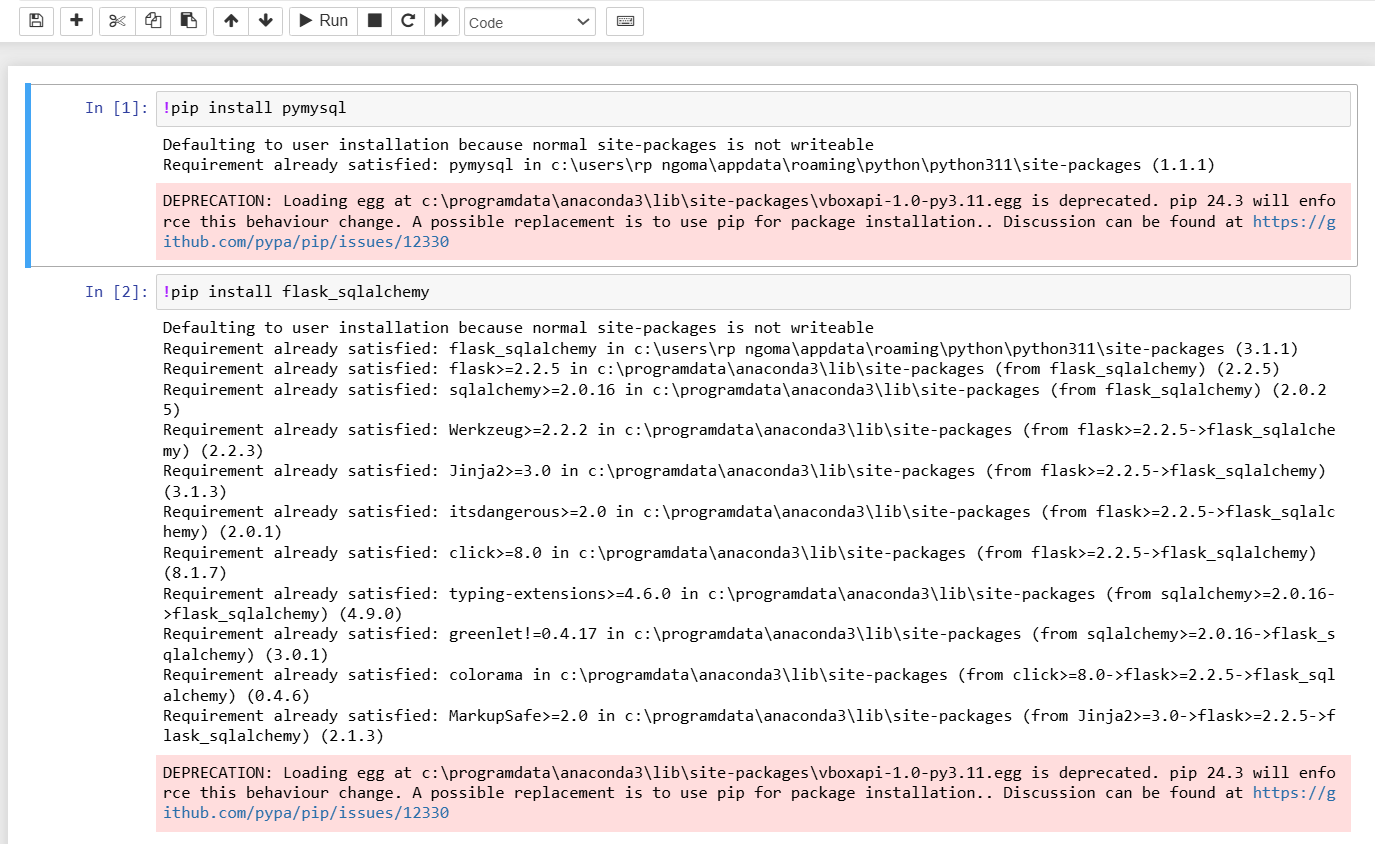
*import pyodbc*

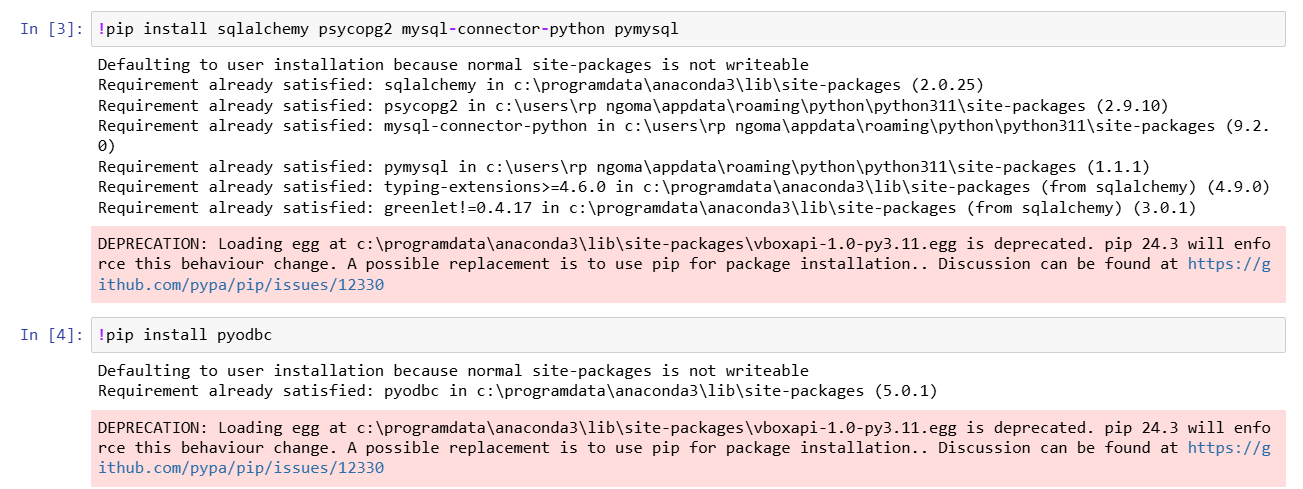
*import pandas as pd*

*from datetime import datetime*

*import warnings*

*warnings.filterwarnings("ignore", category=UserWarning, message="pandas only supports SQLAlchemy connectable")*

****

****

**Database Connection**

Firstly we have been create database by using MySQL I we fulfil records as given in our task. After we extract to the tables and transform these records, then load to the Data warehouse called “gravity\_books\_dwh”.

In order to start our analysis into our warehouse, we firstly connect our databases;

*import mysql.connector*

*import pandas as pd*

*# MySQL connection details*

*db\_config = {*

*"host": "localhost",*

*"user": "root",*

*"password": "",*

*}*

*# Establish MySQL connection*

*conn = mysql.connector.connect(\*\*db\_config)*

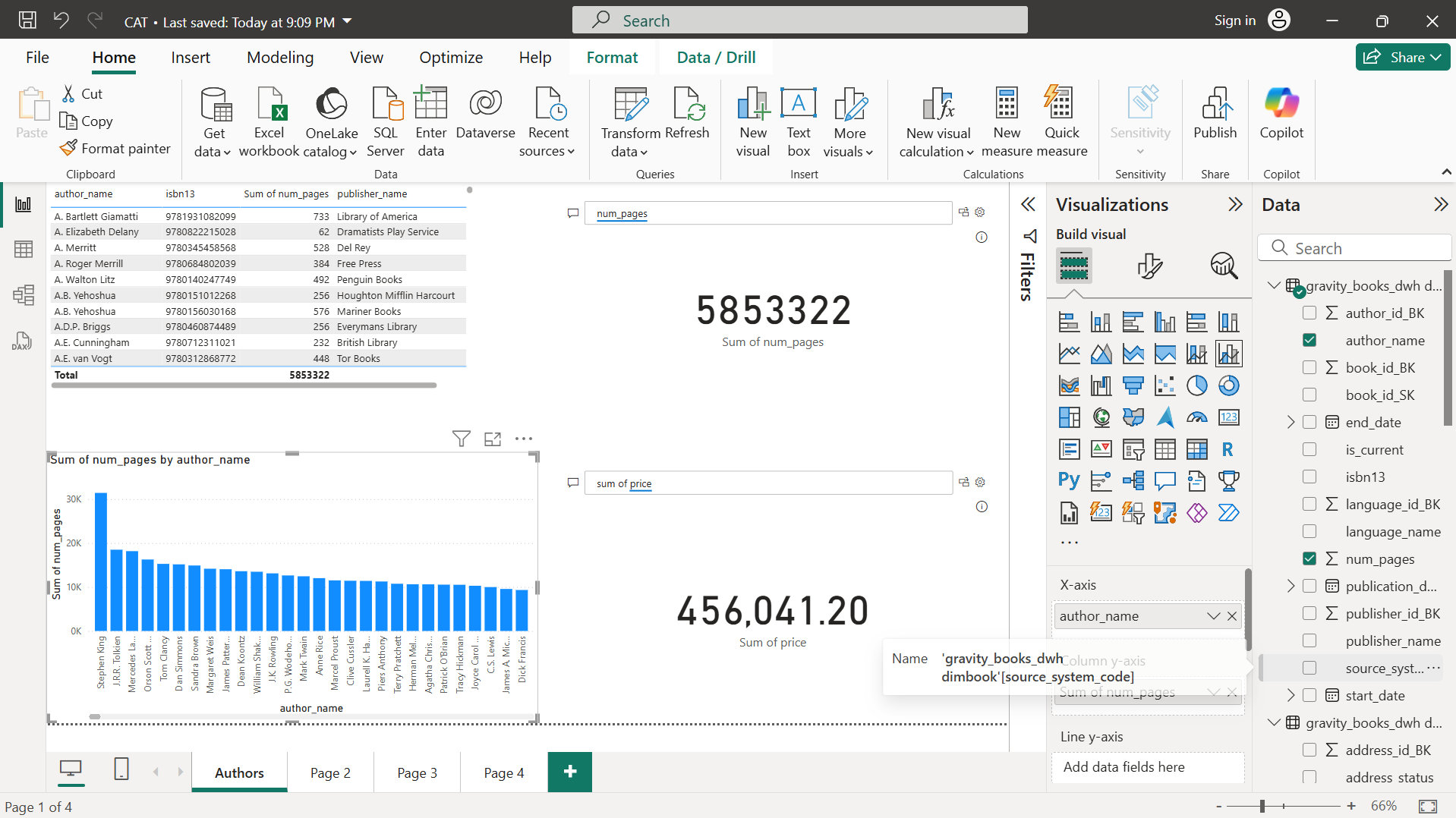
*cursor = conn.cursor()*

*# Step 1: Create the Data Warehouse Database (if not exists)*

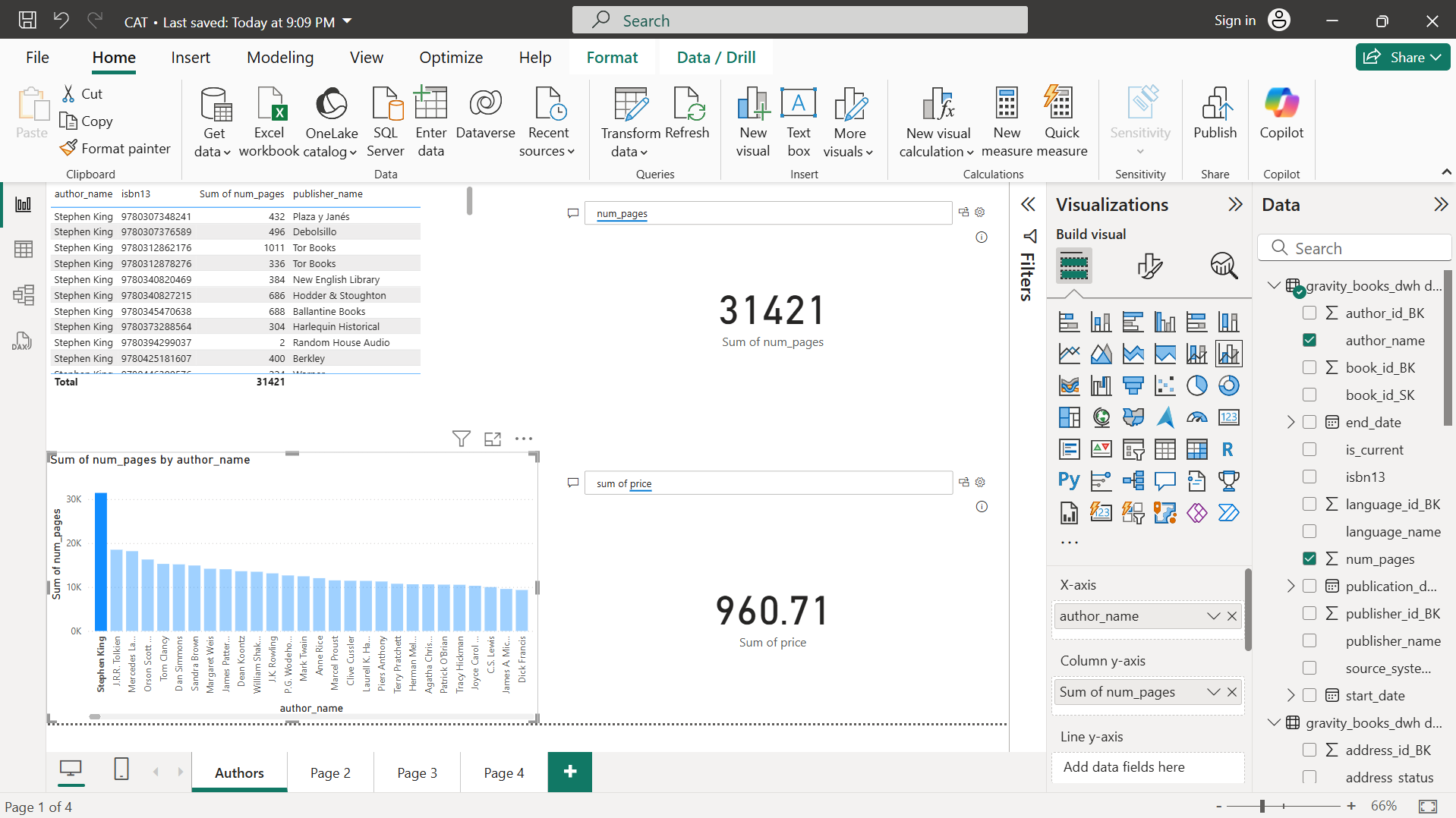
*cursor.execute("CREATE DATABASE IF NOT EXISTS gravity\_books\_DWH;")*

*cursor.execute("USE gravity\_books\_DWH;") # Switch to the new database*

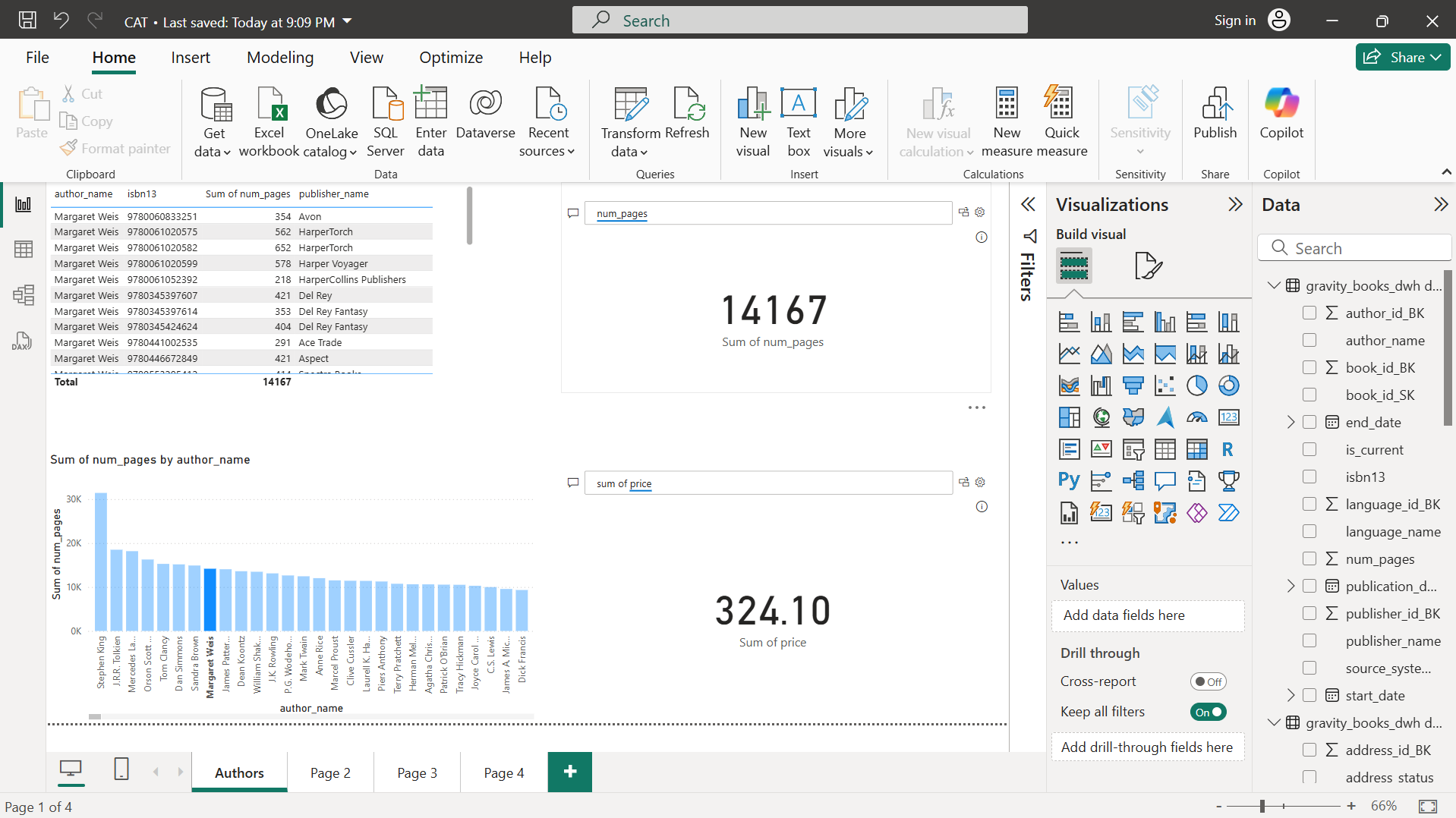
**For starting our Analysis we start on the author**

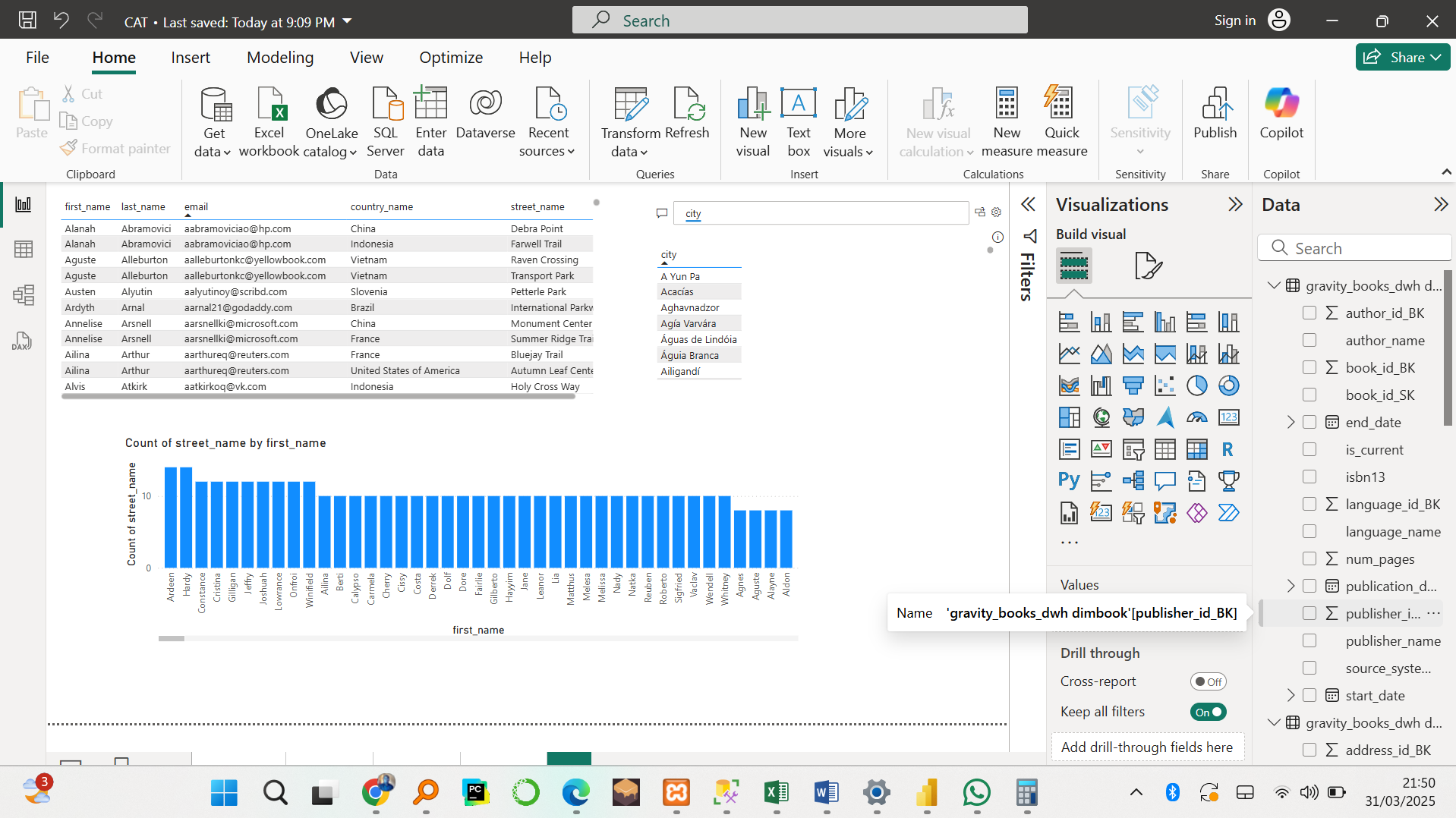
****

In our analysis, we deal with/analyse the author and books he/ she write , this books contained by isbn , sum of number of pages and publisher names which help these author to publish their books. Here as shown on the figure above , is the whole total of pages in our warehouse and the price of all books we have in our ware house. We can not display total of whole warehouse only we can also choose any author as show below.

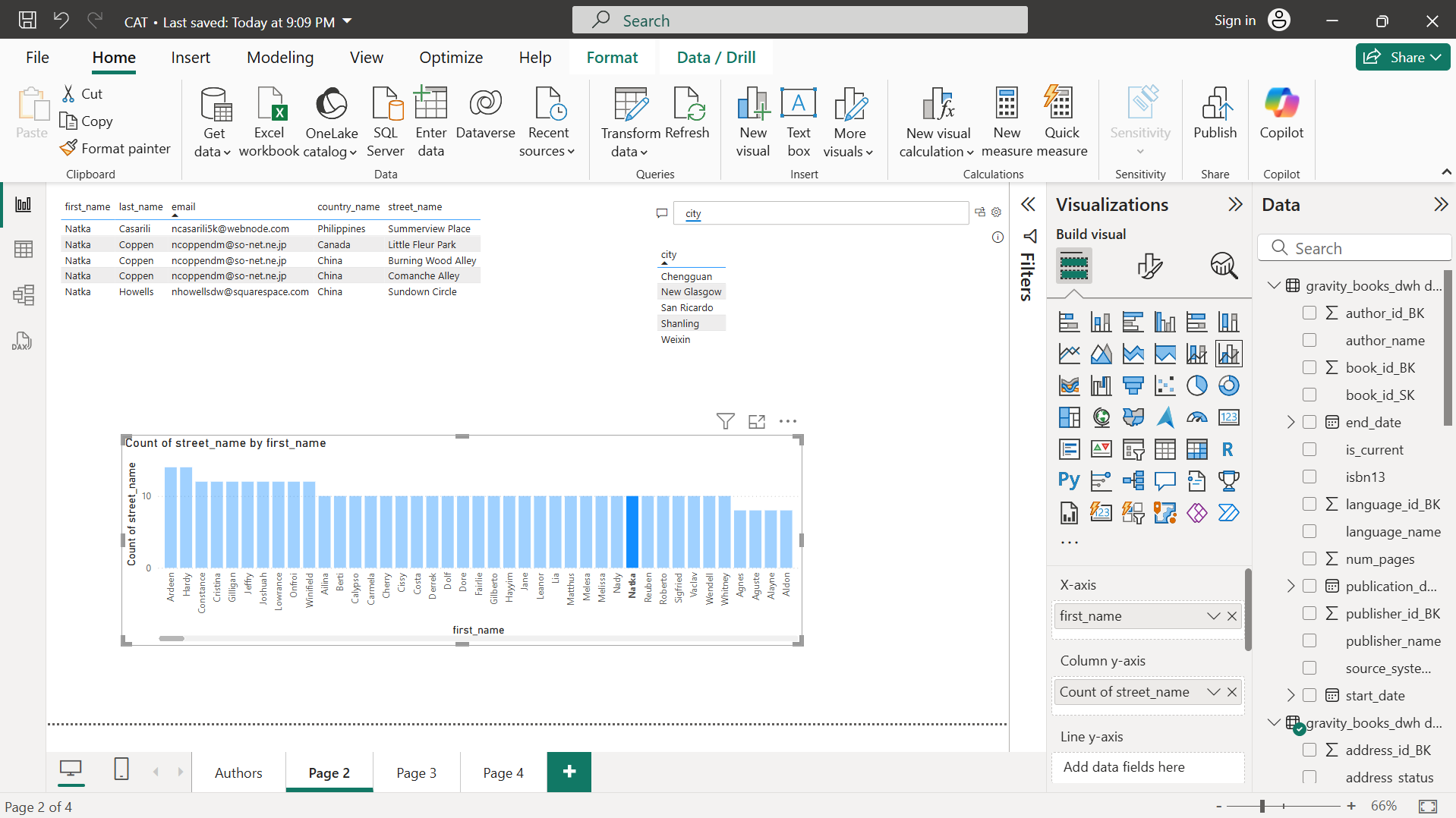


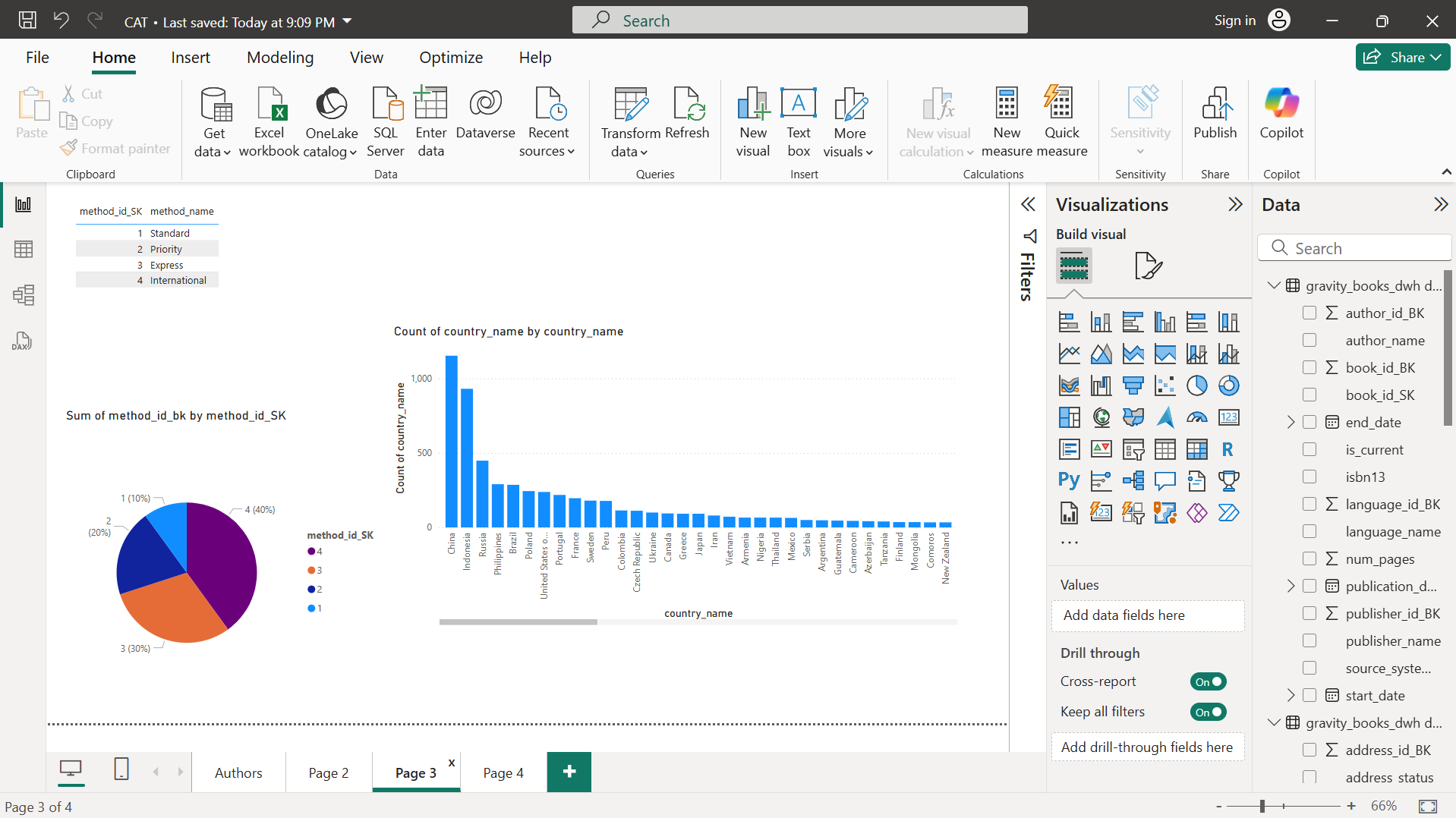
Here we choose one author called “Stephen King”. After choosing him warehouse show total pages of his books and the price of his books.

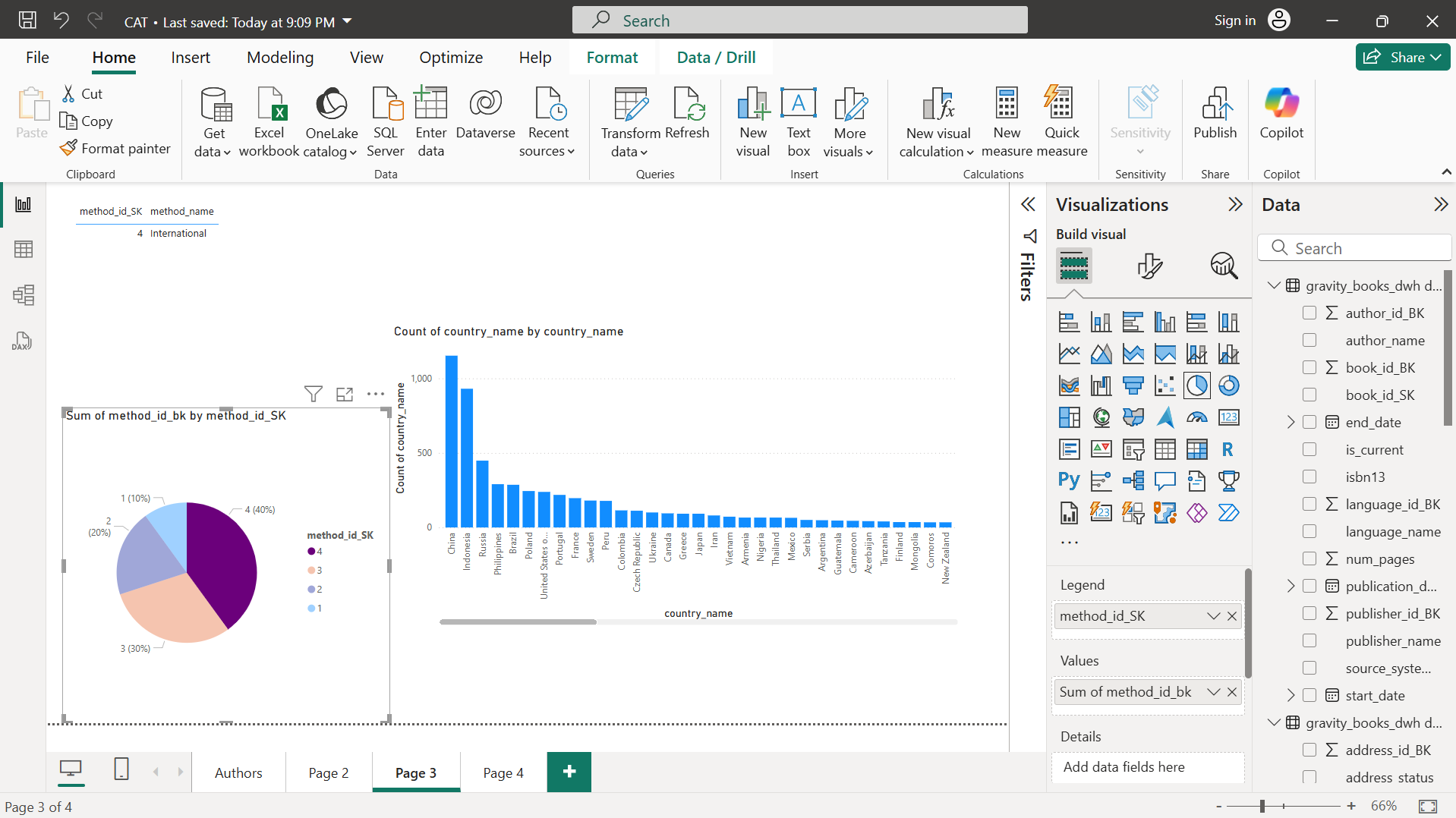


****

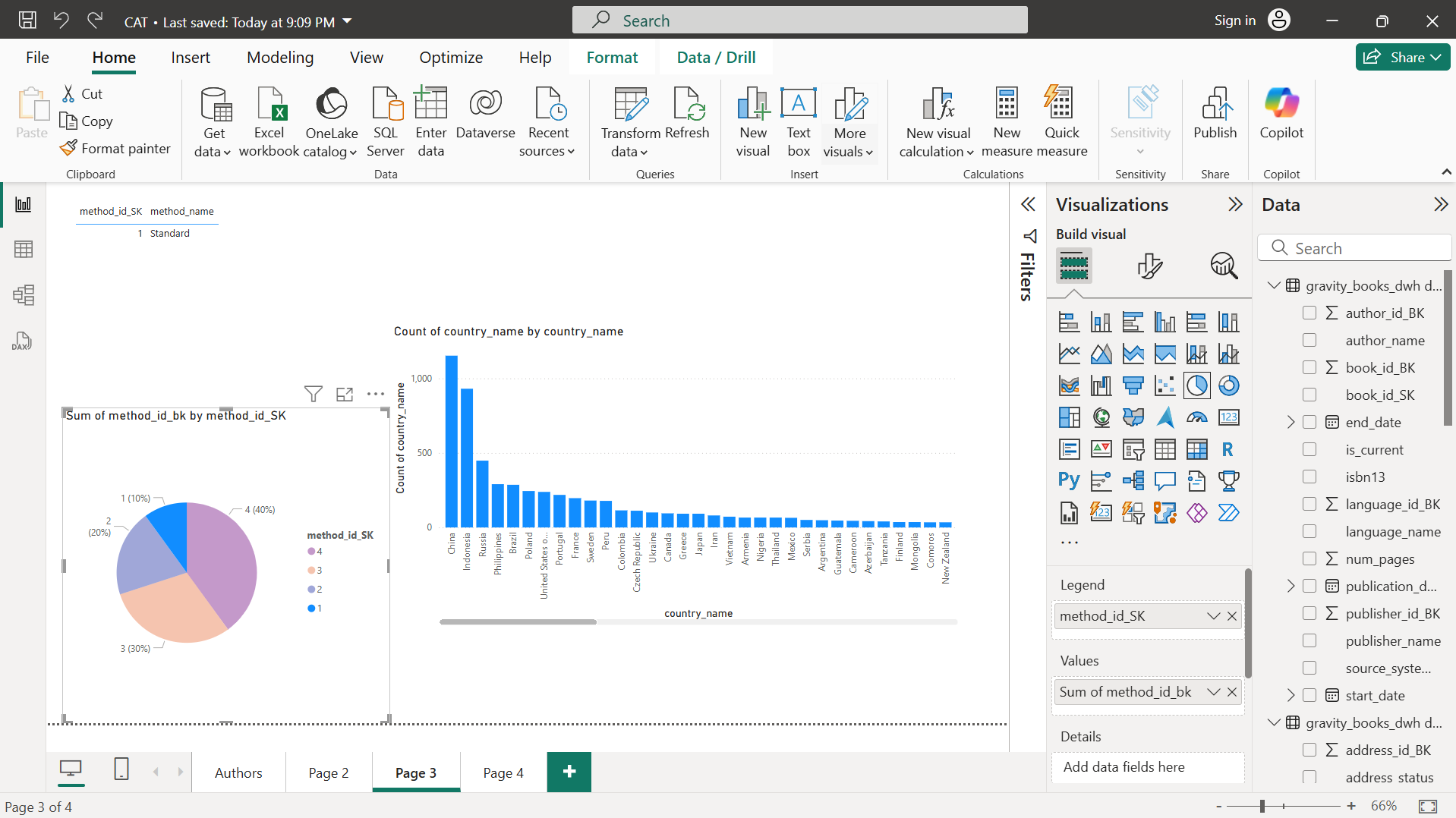
On this picture above, we analyse our customer and their address in order to make sure and know where our customers come from cities, country. we select first name as seen on the dashboard , many customers can have the same first name but directly it display the last name too that help analyst to differ them.



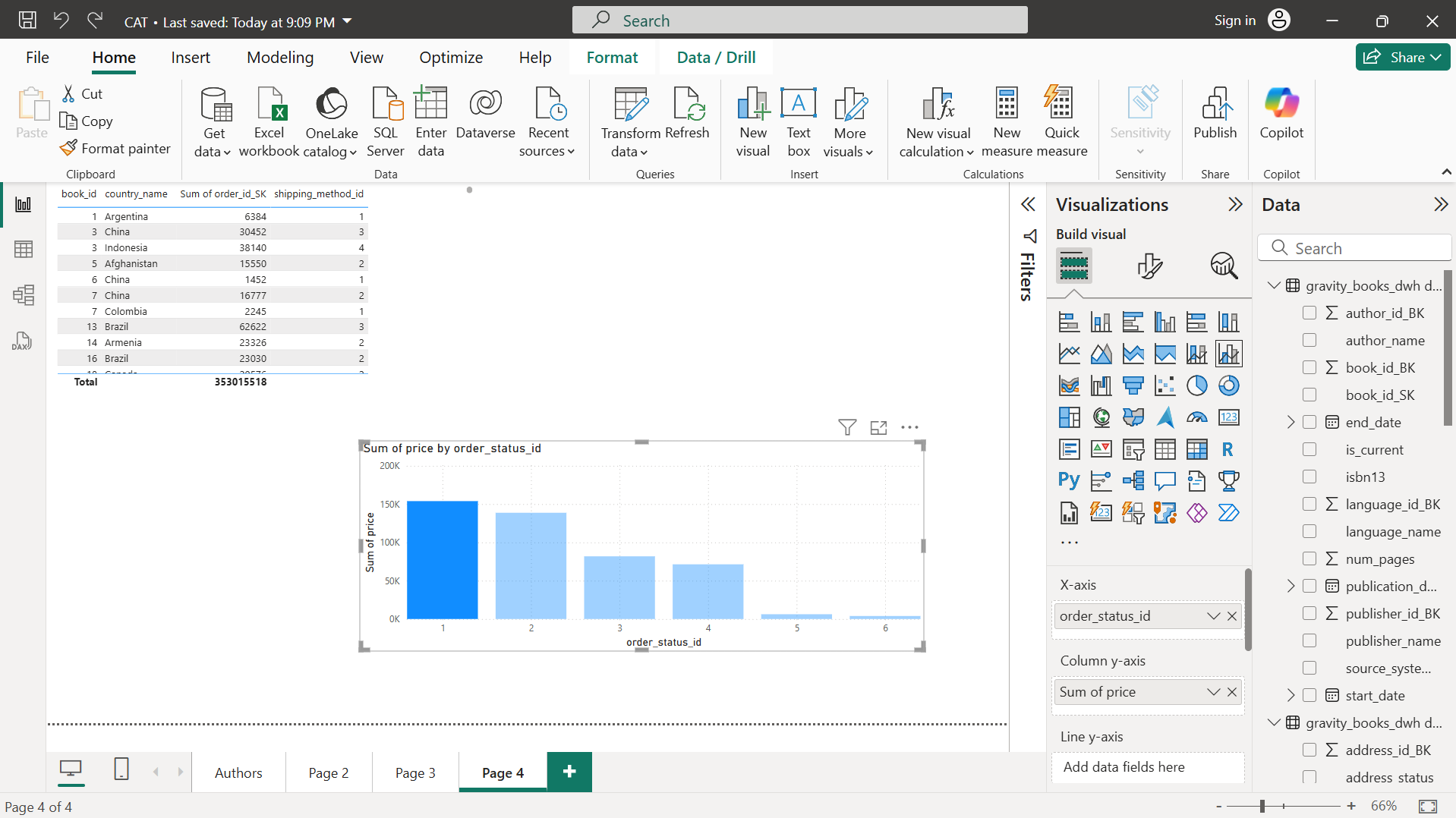
****

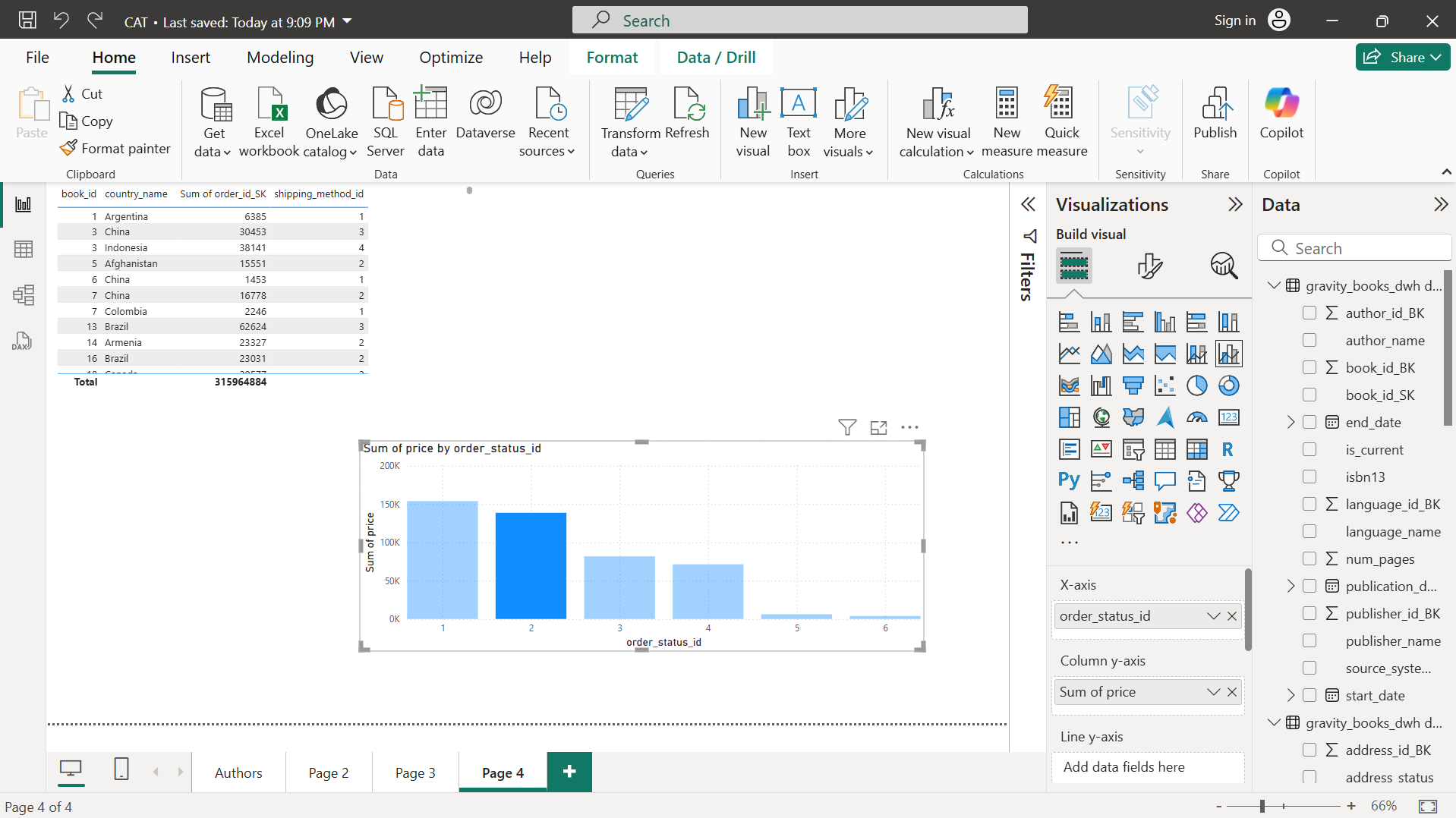


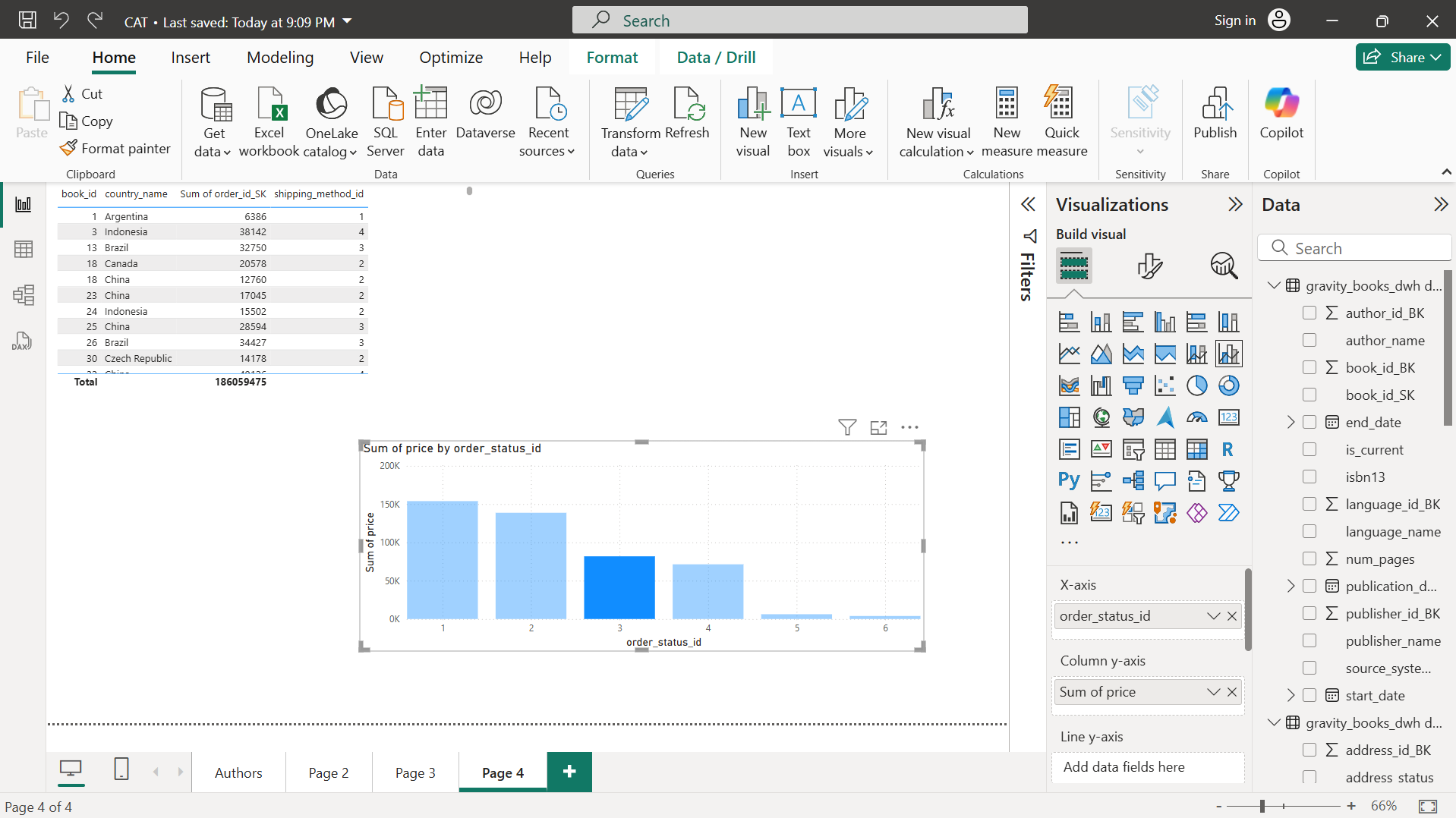
This picture above show analysis of shipping and the methods used and also shows the country that ship more items.



On this dashboard below shows the prices of shipping and the countries who ship accordingly.



**

**